Technical Report 1058

Estimating AFQT by Telephone Using a Computer Adaptive Test

Peter J. Legree, M. A. Fischl, and Paul A. Gade
U.S. Army Research Institute

Michael Wilson
Westat Incorporated

March 1997

19970908 056

United States Army Research Institute
for the Behavioral and Social Sciences

DTIC QUALITY INSPECTED 4

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

EDGAR M. JOHNSON
Director

Technical review by

Robert Kilcullen
Michael G. Rumsey

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1997, March	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) September 1995-December 1996	
4. TITLE AND SUBTITLE Estimating AFQT by Telephone Using a Computer Adaptive Test		5a. CONTRACT OR GRANT NUMBER	
		5b. PROGRAM ELEMENT NUMBER 0603007A	
6. AUTHOR(S) Peter J. Legree, M. A. Fischl, and Paul A. Gade (ARI); and Michael Wilson (Westat)		5c. PROJECT NUMBER A792	
		5d. TASK NUMBER 7501	
		5e. WORK UNIT NUMBER H01	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-RP 5001 Eisenhower Avenue Alexandria, VA 22333-5600		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI	
		11. MONITOR REPORT NUMBER Technical Report 1058	

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

   A computer adaptive test was administered over the telephone by reading items and response alternatives to 144 individuals who had recently enlisted in the U.S. Army and had completed the Armed Services Vocational Aptitude Battery (ASVAB). Subject responses were entered into a computer by the telephone interviewer, thereby allowing the adaptive test program to estimate aptitude with approximately 10 verbal items. Analyses indicate that the Telephone Test is highly correlated with the Armed Forces Qualification Test (AFQT) in the sample we tested, $r = .66$; the bivariate correction for range restriction estimated this population correlation to be .81. A confirmatory factor analysis produced a four factor solution with the Telephone Test loading at a very high level (.91) on a Verbal factor, which had a substantial loading (.72) on a higher order factor. The magnitude of the factor loadings and the administration time (5 to 10 minutes) indicate that the procedure provides an excellent measure of crystallized Verbal aptitude that can be incorporated into brief telephone interviews and used to estimate AFQT and general aptitude.

15. SUBJECT TERMS

Computer adaptive testing   Telephone survey   AFQT estimation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	21	Paul Gade 703/617-8866

i

Technical Report 1058

Estimating AFQT by Telephone Using a Computer Adaptive Test

Peter J. Legree, M. A. Fischl, and Paul A. Gade
U.S. Army Research Institute

Michael Wilson
Westat Incorporated

Organization and Personnel Resources Research Unit
Paul A. Gade, Chief
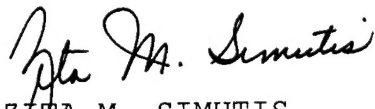
In response to direction provided by the U.S. Army Deputy Chief of Staff for Personnel, the U.S. Army Research Institute is conducting research to evaluate and improve military recruiting. This report describes and evaluates a procedure to test mental quality in telephone surveys. This procedure is important because the military maintains mental quality enlistment standards that screen a substantial proportion of the youth population. Prior to this project, it was not possible to test the mental quality of survey participants and the utility of survey data were therefore limited. This testing procedure will allow survey data to be analyzed by mental quality to identify factors that can be used to encourage the enlistment of high mental quality youth. This capability is important to the military recruiting community for the purposes of analyzing recruiting survey data and formulating recruitment policy.

ZITA M. SIMUTIS
Technical Director

EDGAR M. JOHNSON
Director

ESTIMATING AFQT BY TELEPHONE USING A COMPUTER ADAPTIVE TEST

EXECUTIVE SUMMARY

Research Requirement:

The Department of Defense supports annual surveys of American youth to formulate recruiting policy and advertising campaigns. Although the military maintains minimum mental quality entrance standards, recruiting survey data cannot be segmented by mental quality because no existing procedure can be used to test the mental quality of participants in telephone surveys. This project developed and validated a procedure to test the mental quality of participants during telephone interviews.

Procedure:

A computer adaptive test was administered over the telephone by reading items and response alternatives to 144 recruits who had recently completed the Armed Services Vocational Aptitude Batter (ASVAB). Subject responses were entered into a computer by the telephone interviewer, thereby estimating mental quality with 10 to 12 verbal items in approximately 5 minutes.

Findings:

Analyses indicate that the Telephone Test is highly correlated with the Armed Forces Qualification Test (AFQT) in the sample we tested, $r=.66$; the bivariate correction for range restriction estimated the population correlation to be .81. A confirmatory factor analysis produced a four factor solution with the Telephone Test loading at a very high level (.91) on a Verbal factor, which had a substantial loading (.72) on general aptitude.

Utilization of Findings:

The correlations, the factor loadings, and the administration time estimates indicate that the procedure provides an excellent measure of verbal aptitude that can be incorporated into brief telephone interviews and used to estimate AFQT and general aptitude. A large-scale survey will be conducted during FY97 to further evaluate the Telephone Test and segment the male recruiting market by mental quality for the first time. In addition to its military applications, the Telephone Test may have commercial value because it is the only existing instrument that can provide accurate estimates of mental aptitude in a 5- to 10-minute telephone interview.

ESTIMATING AFQT BY TELEPHONE USING A COMPUTER ADAPTIVE TEST

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Estimating AFQT by Telephone
Using a Computer Adaptive Test[1]

Introduction

Survey research has become increasingly dependent on computer assisted telephone interviewing (CATI) because it is a cost effective procedure to explore the precursors and correlates of a number of important social and practical topics, e.g., criminality, civility, political involvement, career interests and consumer preferences. Although many individuals would concede that these phenomena relate to general cognitive aptitude, large scale telephone surveys have not accurately measured cognitive aptitude or mental quality, and therefore the utility of these survey data has been limited.

The fact that neither cognitive aptitude nor cognitive ability have been accurately tested in telephone interviews may reflect the difficulty of obtaining these data[2]. To our knowledge, only researchers associated with the Colorado Adoption Project (CAP) have accurately tested cognitive abilities over the telephone (Kent & Plomin, 1987; Cardon, Corley, DeFries, Plomin, et al., 1992; DeFries, Plomin, Fulker, 1994). However, the CAP procedure is administratively complex, requires a 35 to 60 minute testing session, and involves the preliminary mailing of a package that is opened during the test administration. The procedure has been used to estimate the cognitive abilities of individuals in small, highly motivated populations, such as the adoptive and biological families followed in twin studies.

---

[1] The computer adaptive test administered in this study was drawn from a slightly larger computer adaptive test that was developed as a preliminary military screening tool. The screening tool is advisory and is not used to make personnel selection decisions, therefore test security should be a minimal concern. This test provides an estimate or acquired verbal and quantitative knowledge and can be characterized as a measure of general aptitude.

2 Crude methods to estimate cognitive ability in telephone interviews have been developed. Luskin (1990) analyzed a database that contains an estimate of mental aptitude, labeled intelligence, based on telephone administrator ratings of survey participants. Luskin also referred to two short vocabulary tests, a 10 item Gallup-Thorndike measure and a 12 item measure attributed to Cattel, that correlated .40. We could not find published information describing these measures.

There are several problems with the CAP procedure that limit its utility for military appplications. First the military selects individuals on the basis of cognitive aptitude as opposed to cognitive ability. Second few telephone survey participants are sufficiently motivated to allow an interview to last longer than 30 minutes. Third it is impossible to mail a package to be used by participants if a random digit dialing procedure is used to select subjects. A final concern is that test materials are difficult to control once mailed. Because of these considerations, the CAP procedure is not suitable for military telephone interview applications. Nonetheless the CAP procedure demonstrates that cognitive tests can be administered over the telephone.

Being able to estimate cognitive aptitude within a telephone interview is particularly important to the U.S. Defense Department because (1) the military conducts an annual telephone survey of American youth in order to track interest in military and non-military careers and formulate recruitment policy, and (2) the military maintains cognitive aptitude entrance requirements. Before this project was initiated, the only means available to the U.S. Defense Department to estimate the cognitive aptitude of surveyed individuals was to employ an algorithm based on demographic variables (Orvis & Gahart, 1989; Stone, Turner & Wiggins, 1992). However, the algorithm provides only marginally accurate estimates of cognitive aptitude, i.e., the algorithm places youth in the top or bottom half of the distribution and is correct for approximately 70 percent of the cases. To the best of our knowledge, this information has not been used to formulate enlistment policy. These considerations highlight the desirability of directly estimating cognitive aptitude with a short test suitable for telephone administration.

The U.S. Army Research Institute developed the Computerized Adaptive Screening Test (CAST) to help U.S. Army recruiters accurately and quickly estimate the cognitive aptitude/mental quality of potential recruits and concentrate their efforts on recruiting individuals who are likely to meet Army entrance requirements. The CAST consists of two types of items, verbal and quantitative. Validity data show that the CAST is highly correlated, .91, with the military's primary estimate of cognitive aptitude, the Armed Forces Qualification Test (AFQT) (Wise, McHenry, Chia, Szenas & McBride, 1989).

The CAST is easily administered. Potential recruits are
seated in front of a computer and answer a series of multiple
choice questions using the keyboard. The CAST selects questions
from a large bank of items, and performance is estimated with
approximately 10 items from the verbal domain and 5 from the
quantitative domain. Administration time is usually about 10
minutes. To determine whether the CAST verbal items could be
administered by telephone and yield an accurate estimate of
general aptitude, we administered the CAST verbal items over the
telephone to 144 individuals for whom operational cognitive
aptitude scores were available.

Method

Armed Services Vocational Aptitude Battery & Telephone Test
Descriptions

The Armed Services Vocational Aptitude Battery (ASVAB) is
the job classification battery used by the military to assign
recruits to military occupations. It is group administered using
scannable answer sheets, requires approximately three hours to
complete and is taken by all military recruits. The ASVAB
consists of 10 multiple choice subtests named for their content
domains: Numerical Operations (NO), Coding Speed (CS), General
Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK),
Paragraph Comprehension (PC), Auto and Shop Information (AS),
Mathematics Knowledge (MK), Mechanical Comprehension (MC), and
Electronics (EL). All ASVAB subtests are power tests except CS
and NO, which are speeded.

Selection into the military is based on the Armed Forces
Qualification Test (AFQT) score, which is calculated by combining
performance on 4 of the 10 ASVAB subtests: AR, MK, PC and WK.
Scores on the other subtests are combined in various ways to
yield composites that are used to make job assignments.

The Telephone Test consists of the verbal items from the
Computer Adaptive Screening Test (CAST). The verbal items are
word knowledge questions requiring recognition of synonyms. The
CAST was developed to predict AFQT. Apart from the deletion of
the CAST quantitative items, no other change was made to the
program.

Participants

Participants were 144 recruits who had recently enlisted
into the military but had not yet attended basic training.
Participant demographics follow: 74 percent male and 26 percent
female; 58 percent Caucasian, 33 percent African American, 4
percent Hispanic, and 6 percent belonging to other groups. The
participants ranged in age from 17 to 32 years with 18 being the
modal age.

Procedure

The U.S. Army provided the social security and telephone
numbers of approximately 300 individuals who had signed

enlistment contracts, but had not yet entered basic training. The social security numbers were used to obtain ASVAB and AFQT scores of the recruits. The Telephone Test administrators were not given access to the ASVAB or AFQT scores.

The Telephone Test scores were obtained by telephoning recruits from an office near Washington, DC. The test administrators were experienced telephone interviewers who were carefully selected to minimize the possible impact of regional differences in accents. After explaining the purpose of the research to the participants, the administrator read each question and response alternative. The recruits indicated their choice for each item, the administrator responded for the participant by entering the choice into a computer at the administrator's station, and the computer proceeded to select the next item. The branching algorithm presented 10 or 11 items to each recruit.

Results

Analyses indicate that the Telephone Test correlates substantially, .66 (p<.001), with the military's primary estimate of cognitive aptitude, the AFQT. Because range restriction due to military entrance requirements substantially attenuates all ASVAB correlations, the bivariate correction for range restriction was used to estimate the correlation between AFQT and the Telephone Test for the general population; this value is .81. The Telephone Test also correlates substantially with most of the ASVAB subtests. Refer to Table 1.

The primary reason the ASVAB subtest scores were obtained was to explore the factor structure of the ASVAB in relation to the Telephone Test. It was hypothesized that the Telephone Test (TT) would load on a Verbal factor because of its content domain, word knowledge. Previous factorings of the ASVAB have described three and four first-order factor solutions (Legree, 1995; Ree & Carretta, 1994; Kass, Mitchell, Grafton & Wing 1983).

The following factor structure was hypothesized and tested using LISREL (Joreskog & Sorbom, 1993): a Verbal factor composed of GS, WK, PC and TT; a Quantitative factor composed of AR and MK; a Technical factor composed of AS, MC, EL and GS; and a Speed factor composed of CS and NO. The first-order factors were hypothesized to load on a general second-order factor, "g".

Based on LISREL modification data, MC was linked to the Quantitative factor. This additional link does not conflict with the main hypothesis being evaluated: does the TT load primarily on the Verbal factor? No other change was incorporated into the model.

The LISREL Goodness of Fit Statistics support the hypothesized factor structure. The Goodness of Fit Index is .938 and the Root Mean Square Residual is .061, indicating that the correlation matrix is adequately reproduced using the hypothesized path diagram and factor loadings. The Chi-square is not significant (p=.059, ns), which means the hypothesized factor structure can not be rejected. The Goodness of Fit statistics are reported in Table 2, the correlation matrix for the latent variables is contained in Table 3, and Figure 1 presents the path diagram and factor loadings.

Table 1. Correlation Matrix.[1]

| | ASVAB Subtests | | | | | | | | | Telephone |
	CS	GS	PC	WK	AR	MK	AS	MC	EL	Test
NO	59	04	23	03	21	26	-03	11	02	06
CS		11	29	17	24	22	-03	16	02	20
GS			37	50	38	39	37	49	47	60
PC				43	30	27	12	23	19	47
WK					28	24	20	27	24	71
AR						69	30	56	36	37
MK							11	47	30	39
AS								51	50	32
MC									51	47
EL										37

Note 1: Decimal Points have been omitted.   Correlations above 17 are significant at the p<.05 level.

Table 2. Goodness of Fit Statistics for the Confirmatory Factor
Analyses.

Statistic	Value
Chi Square Statistic	52.4
Degrees of Freedom	38
Probability	.059
Goodness of Fit Index	.938
Root Mean Square Residual	.061

Table 3. Correlation Matrix of Latent Variables.[1]

	Verbal	Tech	Quant	Speed	g
Verbal	100	49	52	21	72
Tech		100	50	20	68
Quant			100	21	73
Speed				100	29
Psychometric g					100

Note 1: Decimal Points have been omitted.

Figure 1.   Confirmed Path Diagram and Factor Loadings
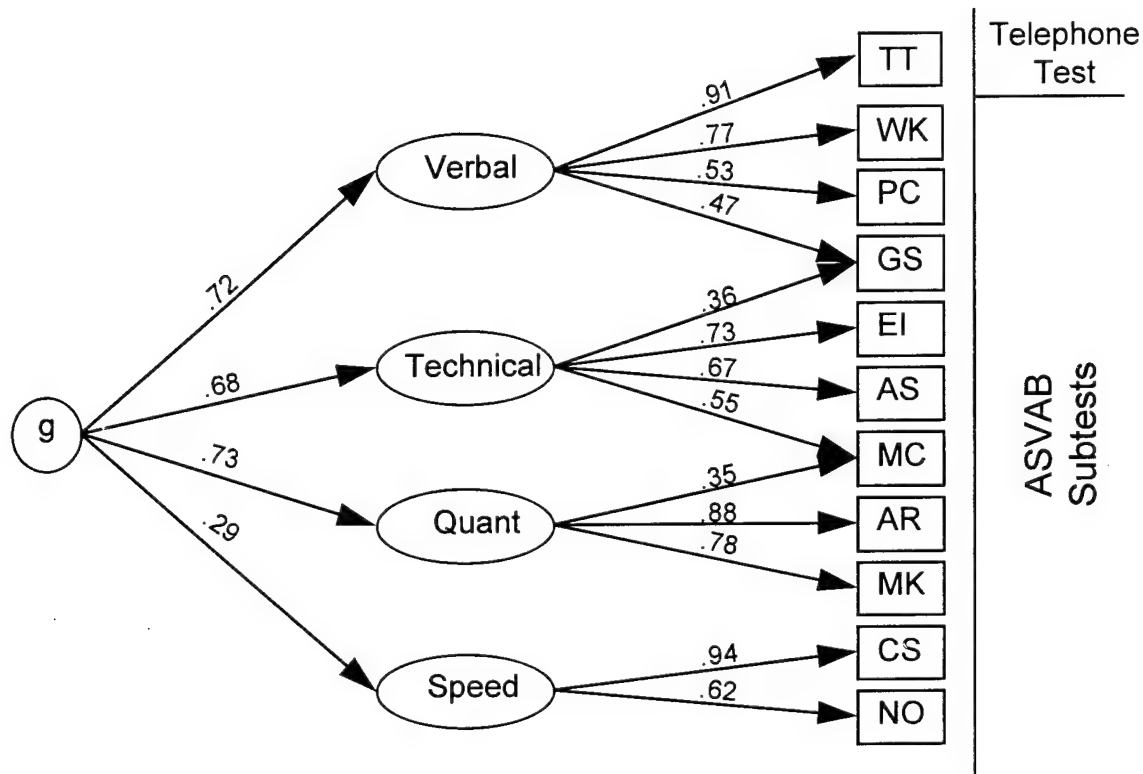
Discussion

The factor and correlational analyses of the Telephone Test
data demonstrate that verbal and general cognitive aptitude can
be accurately estimated using a telephone administration
procedure.   In fact, the loadings of the Telephone Test were
greater on both the Verbal and the General factor than any other
ASVAB subtest.   This is important because the Telephone Test
requires less administration time than any of the ASVAB subtests
and still provides an excellent estimate of verbal aptitude and a
good estimate of general aptitude.

The magnitude of the population correlation estimate between the Telephone Test and AFQT, .81, approaches the population estimate of the correlation between AFQT and the CAST, .90 (Wise, McHenry, Chia, Szenas & McBride, 1989). This indicates that a 5 to 10 minute test, which is administered over the telephone, can provide a reasonably accurate estimate of AFQT. While some incremental validity could have been obtained by adding quantitative items to the Telephone Test, the Telephone Test appears adequate for the purpose of estimating AFQT in recruiting telephone surveys.

The fact that these data were collected using a computer adaptive test is critical. Previous attempts at telephone estimation using non-adaptive tests either required a substantial amount of time (Kent & Plomin, 1987; Cardon, Corley, DeFries, Plomin, et al., 1992; DeFries, Plomin, Fulker, 1994), or achieved limited validity (cf. Luskin, 1990).

It seems safe to speculate that many telephone-based surveys would have provided much more meaningful data had an estimate of cognitive aptitude been incorporated into their design. Based on these analyses, the U.S. Army Research Institute is incorporating the Telephone Test into a moderately large telephone-based survey addressing issues relating to youth career selection. This new study will be the first large scale telephone survey linking individual differences in cognitive aptitude to other variables. It is expected that this study will allow the youth market to be segmented by mental quality in order to identify variables that can be used to focus military recruiting initiatives on the higher quality segments of the youth recruiting market.

There are several questions that have not yet been answered by the current research. Restriction of range due to selection into the military was addressed statistically, hence it is not known how highly the Telephone Test would correlate with general aptitude in lower aptitude populations. If the administration procedure is not too complex for low aptitude populations, then there is no logical reason to expect the correlation to decrease. In fact, it is more likely that the correlation between general cognitive aptitude and the Telephone Test would be higher in lower aptitude groups because cognitive abilities and aptitudes are known to be more highly correlated in lower scoring groups and would therefore be more highly loaded on a general factor (cf. Detterman & Daniel, 1989; Legree, Pifer & Grafton, 1996; Lynn & Cooper, 1994).

Another issue to be addressed is test fairness. Racial and gender analyses were not conducted due to sample size limitations. Past analyses concluded that "the CAST is a valid predictor of AFQT and the CAST is fair to both blacks and females" (Wise, McHenry, Chia, Szenas & McBride, 1989, p. 74). It would be surprising if a different result were associated with testing medium, i.e., telephone administration. This issue, which may have political and legal implications for some applications, will be addressed in future research.

Another possible concern is the robustness of the Telephone Test to regional dialects of the interviewer. If the Telephone Test becomes widely used, then voice recogntion and production technologies, including synthesizers or recordings, could be employed to simplify and further standardize the administration procedures of the scale.

We expect that Quantitative aptitude could also be measured accurately in a telephone interview, and this capability could be important for some purposes. Although the CAST contains quantitative items, we chose not to administer them because they required complex numerical reasoning skills, appeared too difficult for telephone administration and a floor effect seemed likely. (It is notable that examinees are provided paper and pencils to assist with the CAST quantitative items.) Unfortunately, most quantitative items require complex numerical skills and therefore may be difficult to administer during a telephone interview.

However, it may be possible to develop a bank of knowledge based quantitative items that do not require complex numerical skills. Consider these three items, which were created by the authors to assess math knowledge:

1. The number of degrees in a triangle is:

A) 90   *B) 180**   *C) 270*   *D) 360.*

2. Differentiation is most closely related to:

A) Algebra   *B) Trigonometry*   *C) Geometry*   *D) Calculus**

3. The number of points required to define a line is:

A) 1 *B) 2** *C) 3* *D) 4.*

It seems reasonable that a test composed of items like these would load on a Quantitative factor because these items can be characterized as addressing quantitative knowledge. However, unlike most quantitative questions, these items are strictly knowledge based and do not require numerical reasoning skills. The important point is that a test bank composed of these items might provide a useful estimate of quantitative aptitude that could be administered in a telephone interview. For some purposes, this capability would provide a valuable extension to the measure of verbal aptitude we evaluated and improve the capability of measuring general aptitude and AFQT in a telephone interview.

References

Cardon, L., Corley, R., DeFries J., Plomin, R., et al. (1992).
    Factorial validation of a telephone test battery of specific
    cognitive abilities. Personality and Individual
    Differences, 13, 1047-1050

DeFries, J., Plomin, R., Fulker, D. (1994). Nature and nurture
    during middle childhood. Oxford, England: Blackwell
    Publishers.

Detterman, D. K. & Daniel, M. H. (1989). Correlations of mental
    tests with each other and with cognitive variables are
    highest for low IQ groups, Intelligence, 13, 349-359.

Joreskog, K. G., & Sorbom, D. (1993). LISREL 8: Structural
    Equation Modeling with the SIMPLIS Command Language.
    Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Kass, R. A., Mitchell, K. J., Grafton, F. C. & Wing H. (1983).
    Factorial validity of the Armed Services Vocational Aptitude
    Battery, Forms 8, 9 and 10: 1981 Army applicant sample.
    Education and Psychological Measurement, 43, 1077-1087.

Kent, J., & Plomin, R. (1987). Testing specific cognitive
    abilities by telephone and mail. Intelligence, 11, 391-400.

Legree, P. (1995). Evidence for an oblique social intelligence
    factor established with a Likert based testing procedure.
    Intelligence, 21, 247-266.

Legree, P., Pifer, M., & Grafton, F. (1996). Correlations among
    cognitive abilities are lower for higher ability groups.
    Intelligence, 23, 45-57.

Luskin, R. (1990). Explaining political sophistication.
    Political Behavior, 12, 331-361.

Lynn, R. & Cooper, C. (1994). A secular decline in Spearman's g
    in Japan, Current Psychology Developmental, Learning,
    Personality, Social, 13, 3-9.

Orvis, B., & Gahart, M. (1989). Quality-Based Analysis
    Capability for National Youth Surveys: Development,

Application, and Implications for Policy. Santa Monica, CA:
Rand Corporation.

Ree, M. J., & Carretta, T. (1994). Factor analysis of the
    ASVAB: Confirming a Vernon-like structure, Educational and
    Psychological Measurement, 54, 459-463.

Stone, B., Turner, K., & Wiggins, V. (1992). Population
    Propensity Measurement Model: Final Analysis Report.

Wise, L., McHenry, J., Chia, W., Szenas, P. & McBride, J.
    (1989). Refinement of the Computerized Adaptive Screening
    Test. Washington, DC: American Institutes for Research.